AN EXAMINATION OF MECHANISMS TO COMBINE SPEECH AND SOUND FOR DATA ANALYSIS

Demo paper for the ICAD05 workshop "Combining Speech and Sound in the User Interface"

Dr Tony Stockman

Department of Computer Science,
Queen Mary College,
University of London,
Mile End Road,
London E1 4NS
tonys@dcs.qmul.ac.uk

ABSTRACT

Speech based screen reader technology provides the most common means for visually impaired users to access spreadsheets and databases. Such a spoken word interface can be effective if relatively small amounts of data are involved. As the size of the data set grows, it becomes increasingly difficult to carry out data analysis tasks with speech alone. As part of an ongoing study into the use of data sonification to address this problem, we have introduced a sonification cursor, together with limited sonification of text, as mechanisms for helping to integrate the use of speech and sound.

This paper and associated demonstration explore these mechanisms further, by examining the motivation for the approach, analysing its strengths and weaknesses in the context of use and examining the scope for further work. The paper begins by explaining why speech output is the preferred means by which visually impaired users access many applications. Some recent, relevant developments in speech based screen reader systems are then described. The nature of the problems that arise in the use of spreadsheets and databases as the size of the data set increases are explained. Our approach to the use of sonification in assisting access to spreadsheets is then briefly outlined. The specific mechanisms of relevance to this workshop, the sonification cursor and text sonification mechanisms are then explained, and examples given of their use in support of specific data analysis tasks. The paper concludes with a review of what has been achieved, and some directions for future work.

1. INTRODUCTION

Screen readers [1] are the primary mechanism used by visually impaired computer users to access computer systems. Although screen readers exist for other systems such as the Macintosh and Unix, by far the most widely used screen readers, Jaws For Windows (JFW) [2] and Window-Eyes (WE) [3] run under the Microsoft Windows operating system. Both JFW and WE support output in both braille and speech, but for a number of reasons, most users of screen readers either use speech output mode alone, or use braille only in combination with speech. The reasons for the general preference for speech output can be briefly summarised as follows:

1. The additional hardware required for braille output is relatively expensive and is an additional component to be carried in mobile use.
2. A significant proportion of visually impaired computer users are not fluent in reading braille.

In contrast, synthetic speech represents a relatively inexpensive option, which can often be provided via a built-in sound card, and which can deliver output at easily variable speeds, which are only limited by the users perceptive ability [4].

Research into the use of non-speech sound to convey information, known as sonification, has grown rapidly since the early 1990s [5, 6]. There has however, so far, been little take up of this research in screen reader technology. The release of JFW version 5 [7], in September of 2003, represented the first significant use of non-speech sound in a commercial screen reader. This release enabled the use of non-speech sound to customise the feedback provided to users about the current state of the interface. These customisations, called "behaviours" in JFW, are defined in speech manager files, known as schemes. Examples of the ways in which non-speech sounds may be user defined include: to identify when focus has moved to a particular type of interface widget, to signal the different states of a checkbox, to indicate upper/lower case or degree of indentation, or identify the values of HTML attributes. The inclusion of this limited support for the use of non-speech sound is intended to reduce the time and improve the accuracy of user tasks such as navigating between objects in the Windows GUI, and the screen-based proof reading of lengthy documents.

Recent releases of JFW and WE have included improved support for the use of specific database and spreadsheet applications. Examples of these improvements include the ability to provide a summary of the objects and relationships in a database, accessing controls on forms, manipulating relationships between tables, speaking row/column headers and totals, speaking formulas and monitoring cells.

2. PROBLEMS IN SPEECH-BASED ANALYSIS OF DATA

Well designed screen reader systems can provide a good level of access to many applications. Data intensive applications however, have some requirements which conflict with some of the inherent

characteristics in speech based access. Examples of tasks which expose these conflicts are as follows:

1. Obtaining an overview of more than a small number of values. The more data items involved, the harder it becomes to remember the set of values while continuing to navigate through the data space.
2. Comparing data values from different parts of the system, for example, between ables in a database, or across multiple rows/columns or worksheets in a spreadsheet. Difficulties arise here due to the fact that access through speech is very much centered around the location of the cursor. Most screen readers provide support for a reading cursor in addition to the usual PC cursor, which may help to address this problem where the number of data items to be compared is small. As the number of data items to be analysed increases, the user rapidly hits the high cognitive load problem described in point 1 above, but now compounded by the possible need to swap between cursors and the requirement to remember relationships between two sets of data.
3. Navigation to a point in the data space not near a specific boundary. The combination of operating system keystrokes and JFW or WE support enable the easy navigation to the beginning or end of most data objects, but navigation to a specific point within a spreadsheet or table is time consuming if the data value is not unique and the coordinates of the data item are not known.

3. SONIFICATION OF SPREADSHEETS

We have developed an exploratory system to examine the extent to which sonification can improve the accessibility of spreadsheets [8]. The approach we have used is founded on the work of Mansur [9] on auditory graphs, and Brewster [10] on the sonification of tables. Mansur devised a method for line graph sonification called Sound Graphs where the y-axis of the graph is mapped to pitch and the x-axis to time. Movement along the x-axis in time causes notes of different pitches to be played where the frequency of each note is determined by the y value of the graph at that time. Mansur found that after a small amount of training, test subjects were able to identify the overall qualities of the data, such as linearity, monotonicity,and symmetry, on 79-95% of the trials [9].

Brewster and Browne [10, 11, 12] conducted a number of experiments sonifying graphs containing two data series and showed that sonification allowed users to visualise graphs containing two data series while listening to them and assigning different instruments to each range. Brewster and others [13] conducted experiments exploring 2D tables with speech and non-speech sound and discovered that users found the use of pitch to be valuable in determining the shape of the data within the table.

There have already been applications created that allow for the sonification of Excel spreadsheet data. Sonification Sandbox [14] is a flexible tool which supports experimentation using a wide range of sound parameters. From the point of interactivity however, it suffers from the fact that data input is achieved through a comma-delimited, numeric-only file, which is created by extracting row and column subsets from a spreadsheet. Whilst the application allows considerable flexibility in exploring different parameter mappings, it does not afford interactive use through screen readers such as JFW or WE.

The originality of our approach is in the mechanisms provided to support interactive sonification of spreadsheet data, including automatic sonification of arbitrary cell ranges, recalibration of the frequency range for sonifying subsets of a spreadsheet, sonification of composite cell ranges and multiple range sonification.

The system we have developed is intended for use in conjunction with a screen reader, which, for the reasons described in the introduction, is most likely to use speech. It was important then to take into consideration the way in which speech from the screen reader and sounds from the sonification system could be integrated most effectively to support data analysis tasks. The two mechanisms which address this problem directly in our current system are the sonification cursor and the sonification of text.

4. MECHANISMS TO SUPPORT SPEECH AND SOUND INTEGRATION

4.1. Summary of the Use of Cursors in JAWS

JAWS provides three cursors for screen navigation, the PC cursor, the JAWS cursor and the Virtual PC cursor.

The PC cursor is the primary cursor used in JAWS. It is used for most standard Windows system functions and corresponds to the system focus. The PC Cursor is used to navigate menus, edit text in documents and to navigate in dialog boxes. The PC cursor is active when JAWS is first started, but if you switch to another cursor, you can press NUM PAD PLUS to make the PC cursor active once again. The PC cursor can only go where the system focus can go, and this does limit the access it provides to the Windows environment.

The JAWS cursor corresponds to the system mouse. It is used to manipulate the mouse pointer, read static text to which the PC cursor does not have access and to access other parts of Windows to which the PC cursor cannot be moved. When using the JAWS cursor, the system provides keystrokes to simulate clicking of the mouse buttons.

Text in a web page is very much like the static text in a Windows status message. The PC cursor cannot move to this text as there is no insertion point. The JAWS Cursor can access this text, but this is not the most effective way to access larger documents such as those often found on the World Wide Web. JAWS provides the Virtual PC Cursor to simulate an insertion point for use with recent versions of Internet Explorer and related programs. This provides similar functionality to that found in any word processing document. It is possible to read text by word, line, sentence, or paragraph, or select and copy text to the Windows clipboard. The Virtual Cursor is active by default in applications that support it. If you switch to another cursor, you can press NUM PAD PLUS to make the Virtual Cursor active again.

4.2. Use of Cursors to Support Speech and Sound Integration

The sonification cursor, which is used in addition to the PC cursor, enables visually impaired users to navigate spreadsheets employing both speech and non-speech sound. Users initiate the use of the sonification cursor by switching manual sonification to *on*. In this mode, in addition to navigating the spreadsheet with the PC cursor, the sonification cursor can be used to sonify individual cell values as the user navigates the screen. The keys used to navigate the sonification cursor are `S`, `D`, `F` and `E`, forming the same shape as the usual PC cursor keys, allowing the navigation of the

sonification and PC cursors to be carried out with the left and right hands respectively. The manual sonification facilitated by the sonification cursor provides a finer grained control of cell sonification, compared with the various options provided for automatic sonification of specified cell ranges. Users can rapidly navigate spreadsheets manually using the sonification cursor, gaining a feel for the approximate value of numeric cell values. When users find values that they wish to know exactly, they can employ the PC cursor to make the screen reader speak them (the default situation being that the PC cursor is locked to the sonification cursor). If there are values of interest that appear in the same row or column, or in otherwise close to sonified values, the PC cursor can, with a single keypress be unlocked from the sonification cursor, the data values of interest can be checked, and then if desired the PC cursor can be"pushed" back to the sonification cursor and re-locked to it.

Conversely, the exploration of the spreadsheet might be based on speech-based navigation using the PC cursor driving a screen reader, and when required the sonification cursor may be unlocked from this, used to sonify a set of related cells, and then returned and re-locked if required.

For example, a user might navigate using the PC cursor between column headings corresponding to annual sales. When required, the sonification cursor can be unlocked from the PC cursor to provide a rapid overview of variations in monthly sales values appearing in the corresponding column. The sonification cursor can then be re-locked to the PC cursor to navigate to the column for the following year.

While in sonification mode, users can use sonification to assist the rapid identification of up to 3 text strings. The user specifies which text strings are of interest, and selects a pre-recorded sound to be associated with each string.

5. APPLICATION IN DATA ANALYSIS

The spreadsheet sonification system, including the sonification cursor and text sonification, has been used in detailed but informal evaluations with over 20 visually impaired users. The sonification cursor has proved useful when it is required to examine the values of a number of cells quickly, while maintaining the ability to change the direction of the sonification or interrupt it to perform a related task. For example, the monthly sales of products might be examined using the sonification cursor, and, when a particularly high or low value is found, the PC cursor may be unlocked to check the name of the product, before it is snapped back to the sonification cursor. This is particularly useful if the cell containing the sales values is not near the beginning or end of a row, as described in point 3 of Section 2.

Another common situation where the sonification cursor has proved useful is where the user needs to know the exact value of a column or row header, accessed via screen reader speech, followed by an overview of the cell values for that column/row, obtained through use of manual sonification, before snapping the sonification cursor back to the PC cursor typically to examine the next column/row.

Further situations where manual sonification has proved useful include:

- Identification of cells containing incorrect or missing values.
- Navigation to the location of a distinctive cell value where the exact value is unknown.
- Identification of cells with values above or below specific threshold values (in conjunction with the "alarms" feature described in [8]).

The implementation of text sonification is at an early stage, but with the existing implementation it is clearly demonstrable that location of multiple occurrences of specific text strings is quicker using manual sonification than through speech based navigation with the screen reader. The time required to recognise the sound of each cell being sonified manually being shorter than that required to understand enough of the spoken cell value to know whether or not it corresponds to one of the required text strings.

6. CONCLUSIONS

Recent developments in speech based screen reader technology have brought welcome improvements in the support for database and spreadsheet applications for screen reader users. In spite of these improvements, data analysis tasks using a spoken word interface remain difficult for medium to large data sets. Data sonification can help to address some of the inherent limitations of speech in support of data analysis tasks, but this in turn leads to the problem of the effective integration of speech and sound in the user interface.

Use of a separate sonification cursor to enable cell by cell sonification of cell values is a mechanism that enables the allocation of different elements of data analysis tasks to either speech or sound, according to the strengths of each output mode. The ability to use either cursor independently and snap them together at the location of either cursor when required further supports switching between the use of each output mode. Sonification of text appears to hold some promise in assisting the integration of speech and sonification within the context of data analysis, where it is required to identify one of a number of specific text strings within the data.

Both the sonification cursor and text sonification facilities within our current system could be improved substantially and imperical evaluations of their effectiveness are required. Among the many aspects of future work required are the following:

- Empirical evaluations of both mechanisms to quantify the time savings achievable, including an examination of how this varies as the size and other characteristics of the data set are altered.
- Examination of the usability of snapping the two cursors together with an offset (a requirement a number of users have identified in informal evaluations).
- An investigation of the most effective means of extending support for text sonification.

7. REFERENCES

[1] http://www.rnib.org.uk/technology

[2] http://www.freedom-scientific.com

[3] http://www.gwmicro.com

[4] C. Asakawa, H. Takagi, S. Ino and T. Ifukube, "Maximum listening speeds for the blind," in Proc. of the 9th International Conference on Auditory Display, Boston, Ma, 2003.

[5] G. Kramer, "Auditory Display, Sonification, Audification and Auditory Interfaces", Addison Wesley 1994.

[6] G. Kramer, B. Walker, T. Bonebright, P. Cook, J. Flowers, N. Miner and J. Neuhoff, "The Sonification Report for the National Science Foundation," International Community for Auditory Display (ICAD), 1998.Available online at: http://www.icad.org/websiteV2.0/References/nsf.html.

[7] http://www.freedom-scientific.com/fs_products/-software_jaws5intro.asp.

[8] T. Stockman, C. Frauenberger and G. Hind, "Interactive Sonification of Spreadsheets", in Proc. of the 11th International Conference on Auditory Display, July 2005.

[9] D. L. Mansur, "Graphs in Sound: A Numerical Data Analysis Method for the Blind", Unpublished Masters Thesis, University of California, Davis, USA, 1984.

[10] S. A. Brewster, L. M. Brown, M. Burton, R. Ramloll and B. Riedel, "Design Guidelines For Audio Presentation of Graphs and Tables," in Proc. of the International Conference on Auditory Display, Boston University, Boston, MA, USA, July 2003.

[11] S. A. Brewster and L. M. Brown, "Drawing By Ear: Interpreting Sonified Line Graphs," Proc. of the International Conference on Auditory Display, Boston University, Boston, MA, USA, July 2003.

[12] S. Brewster, L. Brown, R. Ramloll and W. Yu, "Browsing Modes For Exploring Sonified Line Graphs," Proc. of the 16th British HCI Conference London, vol. 2, pp. 2-5, September 2002.

[13] S. Brewster, R. Ramloll, B. Riedel and W. Yu, "Using Non-Speech Sounds to Improve Access to 2D Tabular Numerical Information for Visually Impaired Users," Proc. of the HCI 2001 Conference, 2001.

[14] B. N. Walker and J. T. Coltran, "Sonification Sandbox: A Graphical Toolkit For Auditory Graphs," in Proc. of the International Conference on Auditory Display, July 2003. pp.161-164.